Assessing how visual search entropy and engagement predict performance in a multiple-objects tracking air traffic control task

Lanini-Maggi, Sara ; Ruginski, Ian T ; Shipley, Thomas F ; Hurter, Christophe ; Duchowski, Andrew T ; Briesemeister, Benny B ; Lee, Jihyun ; Fabrikant, Sara I

Abstract: Behavioral performance metrics employed to assess the usability of visual displays are increasingly coupled with eye tracking measures to provide additional insights into the decision-making processes supported by visual displays. Eye tracking metrics can be coupled with users' neural data to investigate how human cognition interplays with emotions during visuo-spatial tasks. To contribute to these efforts, we present results of a study in a realistic air traffic control (ATC) setting with animated ATC displays, where ATC experts and novices were presented with an aircraft movement detection task. We find that higher stationary gaze entropy – which indicates a larger spatial distribution of visual gaze on the display – and expertise result in better response accuracy, and that stationary entropy positively predicts response time even after controlling for animation type and expertise. As a secondary contribution, we found that a single component comprised of engagement, measured by EEG and self-reported judgments, spatial abilities, and gaze entropy predicts task accuracy, but not completion time. We also provide MATLAB open source code for calculating the EEG measures utilized in the study. Our findings suggest designing spatial information displays that adapt their content according to users' affective and cognitive states, especially for emotionally laden usage contexts.

DOI: https://doi.org/10.1016/j.chbr.2021.100127

Originally published at:
Lanini-Maggi, Sara; Ruginski, Ian T; Shipley, Thomas F; Hurter, Christophe; Duchowski, Andrew T; Briesemeister, Benny B; Lee, Jihyun; Fabrikant, Sara I (2021). Assessing how visual search entropy and engagement predict performance in a multiple-objects tracking air traffic control task. Computers in Human Behavior Reports, 4:100127.
DOI: https://doi.org/10.1016/j.chbr.2021.100127

Computers in Human Behavior Reports 4 (2021) 100127

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Assessing how visual search entropy and engagement predict performance in a multiple-objects tracking air traffic control task

Sara Lanini-Maggi [a,*], Ian T. Ruginski [a], Thomas F. Shipley [b], Christophe Hurter [c], Andrew T. Duchowski [d], Benny B. Briesemeister [e,f], Jihyun Lee [g], Sara I. Fabrikant [a]

[a] University of Zurich, Department of Geography / Digital Society Initiative, Zurich, 8057, Switzerland
[b] Department of Psychology, RISC lab (Research in Spatial Cognition Lab), Temple University, Philadelphia, PA, 19122, USA
[c] ENAC, Ecole Nationale de l'Aviation Civile, University of Toulouse, Toulouse, 31055, France
[d] Clemson University, School of Computing, Clemson, SC, 29634, USA
[e] Free University, Department of Psychology, Berlin, 14195, Germany
[f] Neurospective GmbH, Berlin, 13589, Germany
[g] ETH, Bio Engineering Laboratory, Basel, 4058, Switzerland

ARTICLE INFO

Keywords:
Visual search entropy
Eye tracking
Electroencephalography
Air traffic control
Animation
Cross-validation

ABSTRACT

Behavioral performance metrics employed to assess the usability of visual displays are increasingly coupled with eye tracking measures to provide additional insights into the decision-making processes supported by visual displays. Eye tracking metrics can be coupled with users' neural data to investigate how human cognition interplays with emotions during visuo-spatial tasks. To contribute to these efforts, we present results of a study in a realistic air traffic control (ATC) setting with animated ATC displays, where ATC experts and novices were presented with an aircraft movement detection task. We find that higher stationary gaze entropy – which indicates a larger spatial distribution of visual gaze on the display – and expertise result in better response accuracy, and that stationary entropy positively predicts response time even after controlling for animation type and expertise. As a secondary contribution, we found that a single component comprised of engagement, measured by EEG and self-reported judgments, spatial abilities, and gaze entropy predicts task accuracy, but not completion time. We also provide MATLAB open source code for calculating the EEG measures utilized in the study. Our findings suggest designing spatial information displays that adapt their content according to users' affective and cognitive states, especially for emotionally laden usage contexts.

1. Introduction

Understanding how to effectively track multiple objects on an animated display is a problem faced by visualization scientists, the aviation industry, and the general public. Commonly, animated displays are used in research and application domains such as movement data analyses, including air traffic control (ATC), crisis management, surveillance, sport analytics, movement ecology, transportation monitoring and forecasting, and for human health monitoring (Chevalier et al., 2016; Dodge, Weibel, Ahearn, Buchin, & Miller, 2016; Hurter, Conversy, & Vinot, 2009; Klein, van der Zwan, & Telea, 2014). Animated displays serve as decision supports in contexts where situational awareness is critical. In the specific case of aviation, air traffic controllers are tasked with extracting relevant information from dynamic visual displays that will inform their decision on when to make landing and safety-critical decisions for aircraft, e.g., based on the speed and orientation of a specific moving aircraft in a crucial area close to an airport. Typical ATC situational awareness tasks include the prompt and accurate perception and understanding of ongoing movement patterns of spatio-temporal data, as well as the accurate projection of current movements into the near future (Endsley, 1995).

To help decision makers such as air traffic controllers (ATCo) effectively perform situational awareness tasks, animations should be adequately designed by considering design principles and users' individual background. The quality of the animation design and the users' cognitive and affective skills, such as their expertise, mental effort and engagement level, impact visuo-spatial task performance (Fabrikant, Rebich, Montello, Andrienko, & Andrienko, 2008; Fish, 2015; Kriglstein, Pohl, & Stachl, 2012; Lowe & Schnotz, 2008; Maggi, Fabrikant, Imbert,

* Corresponding author.
E-mail address: sara.maggi@geo.uzh.ch (S. Lanini-Maggi).

https://doi.org/10.1016/j.chbr.2021.100127
Received 16 February 2021; Received in revised form 24 July 2021; Accepted 27 July 2021
Available online 29 July 2021
2451-9588/

Abbreviations

3D	Three-dimensional
AOI	Area Of Interest
ATC	Air traffic control
ATCo	Air traffic controller
EEG	Electroencephalography
ERP	Event-related potential
ODC	French Operational Display System (ODS) comet design
PCA	Principal Component Analysis
SEEV	The Salience, Effort, Expectancy, Value model
SSSQ	Short Stress State questionnaire ()
UX	User Experience

& Hurter, 2016; Shipley, Fabrikant, & Lautenschütz, 2013). For example, expertise can strengthen visual thinking and support focused attention on relevant visuo-spatial information.

Though much is known about the role of cognitive states and traits in decision-making with visual displays, less is known about the role of emotions and their interactions with cognitive processes. Emotions are especially relevant to ATC decisions given that a mistake could be fatal and financially costly, potentially putting decision-makers under stress. More specifically, higher engagement and motivation can improve cognitive processes such as attentional focus and human-computer interactions (Berka et al., 2007). Contrarily, affective states characterized by low engagement and high distress might negatively impact cognitive performance (Derryberry & Tucker, 1994; Easterbrook, 1959; Harmon-Jones, Gable, & Price, 2012; Roth, 2013), especially vigilance (Kamzanova, Kustubayeva, & Matthews, 2014). In addition, ATCo exhibiting a low level of engagement combined with high cognitive load during ATC tasks, even at high performances, may reflect a lower ability to allocate cognitive resources and greater difficulty in mastering critical situations (Wickens & Tsang, 2015). Knowing about viewers' cognitive and emotional states is important for visual display designers to develop not only aesthetically pleasing, but also understandable and more engaging animated displays, such as emotion sensitive assistance systems for ATCo (Pfeiffer, Valtin, Müller, & Rosenthal, 2016). Effective system design in this application context is especially important given that ATC decisions have a high impact on health and safety.

The current study combines multiple measures of user mental and affective states and traits (e.g., expertise and spatial skills) to examine how to support users making decisions with complex animated displays. Specifically, the current study investigates how visual scanning behavior, measured by gaze entropy, predicts air traffic controller performance in an applied context using two types of air traffic displays (i.e., time-stepped and continuous animated ATC displays), with both expert air traffic controllers and non-experts. As a secondary contribution, we also present an exploratory cross-validation approach to identify individual factors, which include engagement, mental effort and spatial skills, influencing viewers' performance.

1.1. Related work

Animations should be an adequate medium to depict real-time spatio-temporal phenomena, reorientation in time and space, as well as the qualitative aspects of motion, compared to static representations (Tversky, Morrison, & Betrancourt, 2002). However, animations are often depicted in a too fast or complex manner to be effectively tracked or adequately understood. They thus do not always conform to the apprehension principle of good graphics, which also states that external representations should be accurately perceived and appropriately conceived (Tversky et al., 2002).

Five factors guide our visual attention to relevant targets in visual search processing: bottom-up guidance by stimulus salience, top-down feature guidance, guidance by scene properties and meaning, previous history of search over timescales, and modulation of search by the relative item value (Wolfe & Horowitz, 2017). Bottom-up visual attention is guided by preattentive, perceptually salient visual cues, such as items, which differ from their surroundings in their attributes. In ATC animated displays, the (relative) motion, size (or length) and orientation of the moving aircraft are particularly important to identify movement speed or direction changes (Hurter et al., 2009). In addition, it seems that animations are effective if they show only a few graphic elements at a time (Robertson, Fernandez, Fisher, Lee, & Stasko, 2008).

In air traffic control contexts specifically, Cavanagh and Alvarez (2005) highlighted the fact that air traffic controllers can track the movement of several aircraft simultaneously by grouping those objects with similar behavior such as similar speed or orientation. Consequently, not only animation design, but also expertise influences viewers' visuo-spatial processing of complex displays. The continuous processing of spatio-temporal information depicted in animations involves complex interrelationships among bottom-up (stimulus-driven) and top-down (previous knowledge driven) mental mechanisms that can increase cognitive load (de Koning & Jarodzka, 2017; Franconeri, Alvarez, & Cavanagh, 2013; Kriz & Hegarty, 2007; Mayer, 2012), which in turn depend on the training level of the viewer (Bunch & Lloyd, 2006).

Eye tracking technologies are increasingly used to assess visual search processes during the viewing of digital images (Brunyé, Drew, Weaver, & Elmore, 2019, December 1), but they are often not sufficient to characterize the cognitive and/or emotional mechanisms involved in the processing of and decision making with animated displays. Gaze transition and stationary entropy can help to inform how individuals switch and distribute their attention between AOIs in a digital display (see sections 2.3.1 and 2.7.2 for more information; Brunyé et al., 2019; van de Merwe, van Dijk, & Zon, 2012). Stationary entropy can be defined as the distribution of gaze amongst multiple AOIs, where high values indicate more equal attention given across areas. On the other hand, transition entropy can be defined as the frequency of switching between AOIs, where higher values indicate more frequent switching between display AOIs and less predictable visual search. Related work on situational awareness in applied contexts suggests that visual search and attentional guidance used in critical decision-making vary in complex ways. For instance, users' gaze transitions might appear more or less homogeneously distributed and more or less frequent depending on the visual search efficiency, expertise level and emotional state of the participant (Jarodzka & Gerjets, 2010). Usually, fewer gaze transitions are inferred to be indicators of efficient and directed search patterns, while higher transition entropy values are interpreted as a more random, non-efficient or exploratory search (Goldberg & Kotval, 1999; Holmqvist, 2011). Moreover, past ATC research demonstrates that anxiety increases stimulus-driven information processing and thus distraction from task-relevant information affecting focused attention and task performance (Eysenck, Derakshan, Santos, & Calvo, 2007, May). This can be individuated by increases in gaze transition entropy (Allsop & Gray, 2014). Similar to what Krejtz et al. (2015) hypothesized with highly curious users, we can speculate that more engaged users will present lower gaze transitions (i.e., more focused and longer concentration on the same AOI) but more homogeneously distributed across AOIs (i.e., higher stationary entropy).

We need additional approaches to better understand users' emotional states, their perceptual and cognitive abilities, and their potential interactions to be able to better explain why a certain display design works and to predict how it will work in similar usage contexts. One approach for improving this understanding is coupling standard behavioral measurements with neurological data and methodological triangulation – here we use the term "triangulate" to refer to multiple variable cross-validation (Bryman, 1984; Duchowski, 2002; Holmqvist, 2011; Schinazi & Thrash, 2018). As quantitative and qualitative methods have their advantages and disadvantages, one data source might be used to disambiguate other data sources that are combined so the weaknesses of one behavioral data source can be complemented with the strengths of another (Holmqvist, 2011). For example, as Hyrskykari, Ovaska, Majaranta, Räihä, and Lehtinen (2008) argued, the attention to a specific visual element in a graphic display for a significant duration can be interpreted in two contradictory ways: (1) the user is attracted by this stimulus and finds it interesting; or, alternatively and worthwhile exploring, (2) the user does not understand the depicted information and is confused or distracted by it. By coupling eye tracking with electroencephalography (EEG) and measures of emotional state (self-report), it might be possible to further disentangle these two opposing interpretations by taking advantage of theories involving emotional and cognitive states as well.

1.2. The present study

Our primary aim in the current study is to identify common variables that may predict task performance for a multiple object tracking ATC task, i.e., the prompt detection of aircraft speed changes depicted on animations. In particular, we intend to investigate how well we can explain and predict task performance from visual scanning behavior, measured by gaze transition and stationary entropy. Further, we investigate the role that training with a specific task or display, animation design, and viewers' personal traits, such as their spatial abilities, mental effort and stress-related emotional states, have on visual search efficiency. To this end, we aim to shed light on the following research questions (RQ):

- **(RQ1)** *How do visual scanning patterns, measured by gaze entropy, predict performance (i.e. task completion time and accuracy) with a multiple object-tracking task using an ATC animated display? Related, do visual scanning patterns contribute to performance independently of expertise and animation design, factors that are known to influence performance?*
  (H1) We expect that individual background and training will influence participants' visual search strategies, due to differences in viewers' cognitive processes, familiarity with the displays and task at hand, and their intrinsic motivation in succeeding on the task. It could be that ATC experts will demonstrate qualitatively different search patterns from novices, which in turn will differently predict performance when tracking multiple objects, as in 3D volumetric image search (Drew et al., 2013), or in aviation (Fox, Merwin, Marsh, McConkie, & Kramer, 1996). Alternatively, experts and novices could share qualitatively similar search patterns, but with differing levels of search efficiency or correlation between search and performance. We thus hypothesize that higher gaze entropy will be associated with superior task performance (higher accuracy and lower completion time), and the association will be even more pronounced for those with high levels of training or familiarity with the animation design type. In other words, the superior task experience of experts will be explained by more efficient visual search patterns and higher top-down visual processing.
- **(RQ2)** *What additional individual factors (mental effort, emotional states, and spatial abilities) underlying visual scanning behavior will predict task performance?* To answer this question, we set the following competing hypotheses:
  (H2a) These individual factors will vary together and triangulate as follows: Task performance will increase with an increase of gaze entropy, coupled with an increase of engagement and spatial skills, and a decrease of mental effort. Respectively, a decline of task performance will be associated with lower spatial skills, an increase of negative emotional states, and a decrease in gaze entropy (i.e., focus on perceptually salient visual elements over task-relevant ones).
  (H2b) Gaze entropy and these individual factors will contribute independently to task performance.

To test the first research question (RQ1), we utilized a multilevel model approach to determine the influence of individual search patterns, as well as of group differences (expertise and animation design) on task performance. To further investigate the role of viewers' backgrounds and internal states (RQ2), we preformed, in this article, a cross-validation approach to triangulate quantitative neuroscience data streams, traditional task performance metrics, and qualitative data collected with standardized questionnaires in an ATC decision-making study. We wished to more deeply assess users' task performance and visual search efficiency with visuo-spatial tasks using real-world ATC scenarios and ATC animated displays (i.e., continuous and time-stepped animations) depicting aircraft movements.

2. Methods

2.1. The example of French ATC animated displays

To make decisions, ATCo use a radar screen, which displays past, current and future aircraft positions. Their tasks are very complex, and they are mainly focused on monitoring aircraft movements, preventing aircraft collisions, and optimizing traffic flow (e.g., fuel consumption, and landing and take-off regulation). ATCo' training is designed to optimize their decisions in complex and time-dependent situations. Radar screens update aircraft positions on the display using time-stepped animations every 4 s. Due to this visual information update, ATCo can retrieve current aircraft position and speed.

Fig. 1 shows the display system currently used by French ATCo at the operational level (Hurter & Conversy, 2008). Aircraft are depicted by multiple squares. The biggest square represents the current position of an aircraft, while the squares of decreasing sizes indicate its past positions. Speeds and accelerations, as well as direction changes, of an aircraft can also be directly inferred from the motion of the aircraft and from the distance (or orientation) between past and current positions of an aircraft (i.e., the faster the aircraft is traveling, the more separated the squares representing past and current positions). Close to the aircraft, ATCo can also visualize and gather a variety of information about the aircraft, such as the speed, altitude, heading, destination, type of aircraft, flight call-signs, and flight intentions. At the same time, they have also to communicate with and give instructions to pilots during landing or take-off maneuvers. A more complete description and investigations of ATC activity can be found in past literature (Cordeil, Dwyer, & Hurter, 2016; Hurter, Lesbordes, Letondal, Vinot, & Conversy, 2012; Letondal, Hurter, Lesbordes, Vinot, & Conversy, 2013; MacKay, 1999; Mackay, Fayard, Frobert, & Medini, 1998) (see Fig. 2).

Due to the high amount of information and tasks that they have to process and accomplish simultaneously, both visually and auditory, ATCo are often under conditions of high cognitive load. For this experiment, however, we decided to focus only on the prompt detection of aircraft speed changes because of our desire to better understand the fundamental aspects of visual attention underlying, as well as the engagement level and the animation design choices best suited to effectively fulfill, basic situational awareness tasks. Specifically, we focused on participant's performance in solving a task involving the first level of situational awareness from Endsley (1995)'s model, i.e., how well and fast participants perceive the visual elements that continuously changes over time on a dynamic scene. The additional two levels have been assessed in a supplementary study; the outcomes are not presented in this publication.

2.2. Experimental design

As shown on Fig. 3, our user study utilized a between-subject design, which included two expertise groups (ATC experts versus ATC novices) and two ATC animation designs (time-stepped versus continuous animations). In total, we tested 16 air traffic animations depicting aircraft movements. The displays were created using Processing (https://process

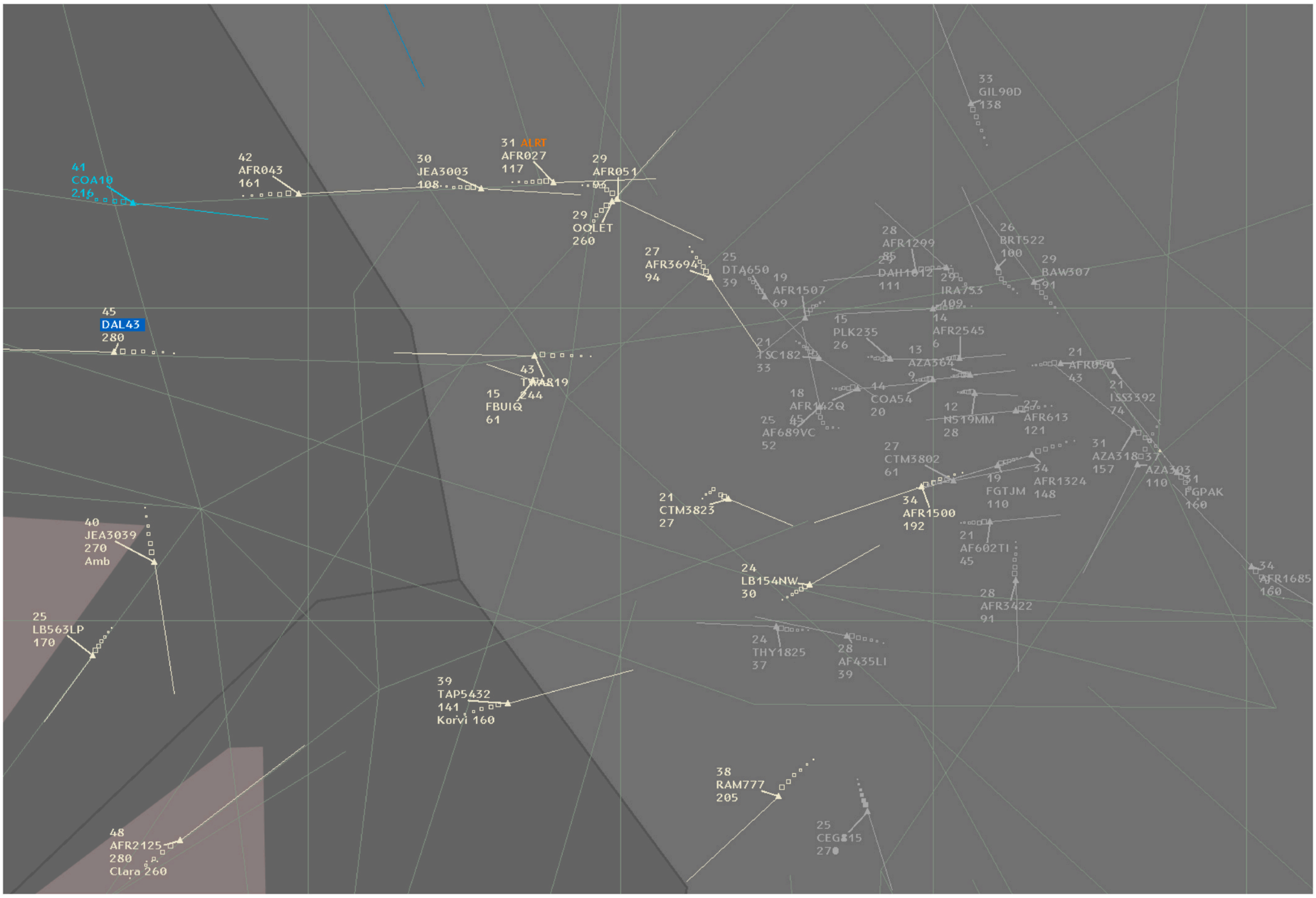

Fig. 1. Screenshot of a French ATC radar screen depicting several moving aircraft.

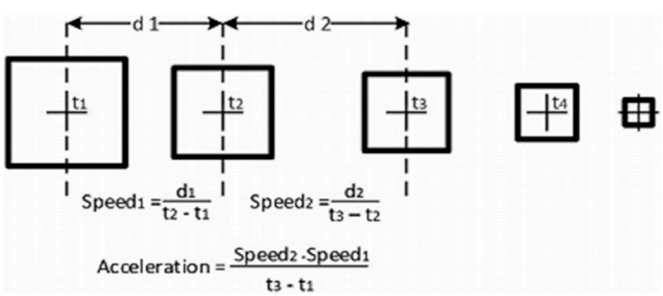

Fig. 2. The representation of one aircraft according to the French ODC comet design (Hurter & Conversy, 2008; Maggi et al., 2016).

ing.org/, accessed: 21.01.2021), and designed-based operational display system displays and realistic ATC scenarios (Fig. 4). We tested participants' task performance (i.e., response accuracy and completion time in the detection of accelerating aircraft) in the two different ATC animation display conditions. More specifically, we manipulated the visual rate of change (smoothness of the transitions between scenes) of the animations (DiBiase, MacEachren, Krygier, & Reeves, 1992). Dynamic visual variables such as rate of change are typically used in cartography to control the visual appearance and dynamics of the animation scenes. Time-stepped animations depict aircraft movements at 4-s intervals, i.e., scenes are refreshed abruptly with one frame every 4 s, the rate currently used at the operational level by French ATCo (Hurter & Conversy, 2008). Conversely, continuous animations display aircraft movements continuously, i.e., updated at 60 frames per second. Such displays are not used professionally for technical and historical reasons, even if the continuous visualization and smooth transitions of the aircraft movements seems to be more coherent with the congruence principle of effective graphics than time-stepped animations. This principle states that the content and format of a visual display should depict the visual information congruently with the mental representations of the user, in this case, real aircraft movements visualized with continuous changes over time (Tversky et al., 2002). Both animation types (including the Processing code) can be visualized at and downloaded from https://osf.io/d2jpg/ (accessed: 21.01.2021).

We manipulated the difficulty of the task by varying the amount of the depicted information and the speed of the moving visual entities. Specifically, eight of the 16 animations depict four moving aircraft (i.e., presenting a lower difficulty level) and eight animations show eight moving aircraft (i.e., presenting a higher difficulty level). The aircraft movements were depicted using typical take-off speeds (i.e., 160, 200, 250 and 290 knots, or kts). Four animations show aircraft moving at the same speed; the other four animations show aircraft moving at different speeds. The speed and heading are kept constant for the whole duration of each animation with the exception of one aircraft, which after 4 s started to accelerate slowly (0.4 kt/s, 1 knot = 0.514 m/s). However, we do not present differences in participants' cognitive effort between these two conditions (4 versus 8 moving objects), as the conditions were selected to provide a range of variability in the stimuli and were not directly relevant to our research questions. The EEG signals recorded within this timeframe were processed to measure participants' cognitive load, frontal alpha asymmetry (FAA), and engagement level. Before every animation, a white screen is shown for 500 ms to the participants to provide a baseline EEG signal.

The task required a visual search task (Posner, 1980) to correctly

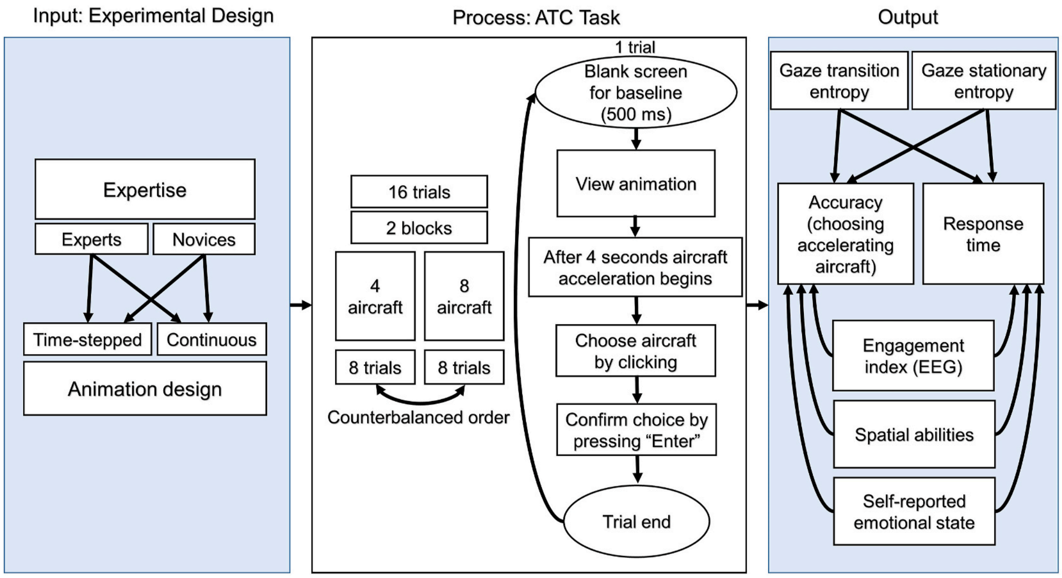

Fig. 3. Input-process-output model of the experiment.

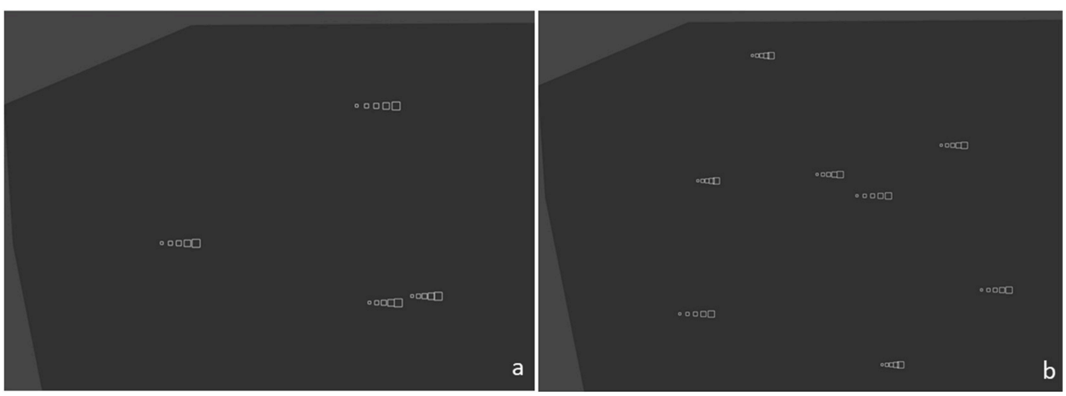

Fig. 4. Video screen capture of 2 test stimuli depicting (a) 4 aircraft moving at four different speeds, and (b) 8 aircraft moving at three different speeds (the whole animations can be seen at: https://osf.io/d2jpg/, accessed: 21.01.2021).

identify the accelerating aircraft – according to the first (perception) situational awareness level (Endsley, 1995) – and to click on the aircraft as quickly as possible. To confirm their choice and move on to the next animation, participants are asked to press the "enter" key on the keyboard. Only one of the depicted aircraft for each trial accelerates (i.e., task-relevant item or target), the remaining features have to be considered as distractors. To successfully complete the task, participants had to avoid distraction by irrelevant but perceptually salient objects, e.g., by fast aircraft movements. To distinguish participants' visual attention between thematically relevant and perceptually salient information, we designed displays so that the accelerating aircraft was never the fastest aircraft.

2.3. Measures

In this study we combined eye tracking, EEG and self-reports data to measure users' cognitive states, traits, and emotional reactions (Westerman, Sutherland, Robinson, Powell, & Tuck, 2007) that might influence visuo-spatial decision-making. We triangulated eye tracking measures (i.e., gaze entropy; see section 2.3.1 for more), self-reported engagement (extracted from the Short Stress State questionnaire; Helton, 2004), and engagement index, alpha power and frontal alpha asymmetry (extracted from EEG data; Coan & Allen, 2004; Heger, Putze, & Schultz, 2010; Kothe & Makeig, 2011; Wolfe & Horowitz, 2017). Table 1 summarizes and defines the physiological and cognitive constructs and the related measures.

2.3.1. Gaze entropy

Entropy measures are used to compare transition matrices, which can indicate whether participants used a directed or a more random search strategy for identifying the accelerating aircraft (Shannon, 1948). Specifically, we used transition entropy to determine how participants' eye movement sequences transitioned between animated AOIs (4 or 8 depending on the number of animated aircraft on a given trial, Krejtz, Duchowski, & Krejtz, 2014) and stationary entropy to determine the distribution of participants eye movements amongst AOIs. More specifically, gaze transition entropy describes "*the rate of fixation transitions between defined spatial regions*", indicating an "*overall estimation for the level of complexity or randomness in the patterns of visual scanning relative to stationary entropy, i.e. the overall spatial dispersion of gaze, where higher entropy suggests less predictability*" (Shiferaw, Downey, & Crewther, 2019, January 1). It is a measure of visual scanning efficiency, where "*the optimal range in gaze transition entropy can be considered the ideal level of scanning complexity that results from modulation of the underlying bottom-up influence (i.e. distribution of salience within the visual field) by top-down prediction (e.g., requirement of the task or the observer's prior knowledge*", and where "*gaze transition entropy is expected to increase with greater top-down engagement*" (Shiferaw et al., 2019). Concurrent analysis

Table 1
Cognitive and physiological sensing methods used in our user study, including the variables measured and the predicted (or analyzed) user cognitive and emotional processes. Further details defining measures and associated data processing are provided in the referenced sections.

Cognitive and sensing method	Measures	Prediction or post-hoc evaluation of the following cognitive and affective states
Neural (EEG)	Electrical activity of the brain using EEG. Measures of interest included alpha power, engagement index, and frontal alpha asymmetry (FAA)	Cognitive load (measured with alpha power, Bunch & Lloyd, 2006a, mental engagement (measured with engagement index, Pope, Bogart, & Bartolome, 1995), and approach versus withdrawal-related motivation (measured with FAA, Briesemeister, Tamm, Heine, & Jacobs, 2013)
Eye tracking	Eye movements and derived measures, e.g., transition and stationary entropy	Search strategies and search efficiency measured with transition and stationary entropy, which differentiate between directed versus randomly distributed gaze transitions over the animation scene and across their graphic elements, (e.g., a focus on specific elements or homogeneous gaze transitions between all the depicted elements) according to Krejtz et al. (2015)
Subjective self-reported	Short stress state questionnaire (SSSQ): Measures subjective emotional engagement, distress and worry	Assessment of self-perceived stress states, i.e., engagement, distress and worry during the task; questionnaire previously validated by Helton (2004)

of both gaze transition and stationary entropy can provide more precise insights into the viewer's visual scanning behavior, as follows (according to Shiferaw et al., 2019):

- an increase of both entropy types may reflect the influence of top-down modulation resulting in a less structured and a greater dispersion of gaze;
- a decrease of transition entropy accompanied by an increase of stationary entropy may suggest distractibility or greater bottom-up influence on gaze control;
- a decrease of both entropy types may indicate a low top-down modulation which results in an insufficient exploration;
- a decrease of stationary entropy when transition entropy is above the optimal range may suggest a top-down interference that leads the viewers to concentrate their eye fixations on only certain elements within the visual scene.

However, how both entropy types are interconnected and predict visual scanning efficiency also depends on the task and visual scene. Gaze transition entropy, as a measure of visual search efficiency, gives insights into top-down modulation of gaze control. Gaze stationary entropy, on the other hand, indicates changes in the spatial dispersion of eye fixations, which can in turn be influenced by viewers' top-down engagement (Shiferaw et al., 2019). More on how gaze transition entropy is calculated can be found in section 2.7.2.

2.4. Participants

A total of 37 participants (*M* age = 30, range from 19 to 45) took part in our experiment. 18 were experts in the domain of ATC (ages ranged from 30 to 45, mean age = 38, standard deviation = 3.53; 3 females and 15 males), with more than 10 years of experience as ATCos in an airport or in a air traffic control center, and as instructors at the "Ecole Nationale de l'Aviation Civile" (ENAC) in Toulouse, France. 19 participants were novices (ages ranged from 19 to 39, mean age = 21, standard deviation = 4.57; 12 females and 7 males), all psychology students at Temple University in Philadelphia (USA), who had no prior knowledge of ATC. Novices received class credits for their participation in the study. Equal numbers of experts and novices participated in the experiment with only one type of animation design. This study was run in accordance with the recommendations of the Temple University's Human Research Protection Program (HRPP) prior to data collection. The ethics procedure was approved by Temple University's Office for Human Subjects Protections, Institutional Review Board (IRB). All participants, both ATC experts and ATC novices, gave written informed consent to participate in this study.

2.5. Apparatus

2.5.1. User experience and self-reported emotional states

Participants' emotional states were assessed by means of the Short Stress State Questionnaire (Helton, 2004) to measure their task engagement, distress, and worry. In total, participants were presented 24 questions to answer before and after the ATC trials. Participants were offered a choice of five responses (1 = not at all, 2 = a little bit, 3 = somewhat, 4 = very much, 5 = extremely). For each participant, the responses of the post-test SSSQ were subtracted by the responses of the pre-test SSSQ, and the result divided by the (individual) standard deviation (σ) according to equation (1) (Helton, 2004), below:

$$\frac{Post - score \; - \; Pre - score}{\sigma} \tag{1}$$

Furthermore, after the experiment, participants were asked to respond to two questions about the task easiness and their enjoyment with the proposed interface and animation design type used to solve the required task. The two questions were:

- *How easy can you detect aircraft movement changes with this kind of display?*
- *Did you enjoy solving the task with this kind of visualization?*

Participants were offered a choice of five responses (1 = Not easy at all/Not at all, 5 = Very easy/Yes, absolutely).

2.5.2. Spatial abilities test

To measure the participants' spatial abilities we used a Hidden Pattern Test (Ekstrom, French, Harman, & Dermen, 1976). This previously validated test measures participant's visuo-perceptual speed, which was especially relevant for the visual detection task in the current study.

2.5.3. Emotiv EPOC + EEG

We measured participants' brain activity with Emotiv EPOC+, a high-resolution 14-channel mobile EEG (Emotiv Inc., https://www.emotiv.com/, accessed: 21.01.2021). The saline-based electrodes for brain activity measurements are located at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4. Following the default setting of the manufacturer, electrodes at P3 and P4 were used as common mode sense (CMS) and driven right leg (DRL) electrodes respectively to provide reference to the EEG measurement. The sampling rate was 128 Hz.

2.5.4. Tobii Studio

Gaze data were collected with the Tobii TX300 eye tracker (Tobii Technology AB, http://www.tobii.com/, accessed: 21.01.2021), and calibrated, processed and analyzed with the software Tobii Studio 3.4. As I-VT fixation filter, we selected a minimum fixation duration of 60 ms, and a velocity threshold of 30°/s. The sampling rate was 300 Hz.

2.6. Procedure

After participants signed the consent form, they were asked to provide demographics, and complete the SSSQ questionnaire. Successively, they were asked to solve the spatial abilities test. Next, the main experiment with animated displays were carried out according to the experimental design described above. At the beginning of the experiment, participants' eye movements and EEG signals were calibrated. After an introductory training with air traffic displays, participants started the main experiment by watching 16 animations and engaging in the required task for each animation, i.e., detecting the accelerating aircraft among four or eight moving aircraft. The animations were presented to the participants in random order on a computer screen with a resolution of 1920 × 1200 pixels. In total, the main experiment lasted about 16 min. Once they completed the main experiment, participants were asked to complete the Short Stress State Questionnaire (Helton, 2004) for a second time, and a post-test questionnaire about their user experience and enjoyment. Finally, participants were thanked for their participation and novices received class credits.

2.7. Data processing

2.7.1. Missing data and data synchronization

Due to low signal quality, 35% of the collected EEG data and 38% of the eye tracking data was missing and not further analyzed to avoid misinterpretation of the outcomes (Dalrymple, Manner, Harmelink, Teska, & Elison, 2018; Wass, Forssman, & Leppänen, 2014, Table 2 for a summary). A complete set of raw data (recorded from all the eye tracking, EEG, and questionnaire data sources) was only available for 19 participants (10 ATC experts and 9 novices). Specifically, the eye tracking data of 13 participants were not further analyzed because the quantity of gaze samples collected by the eye tracker was below a minimum value of 50%, likely due to head or body movements. The 19 remaining participants whose data was included in the multivariate analysis have a valid gaze samples mean of 81%. The EEG data of 13 participants was missing because of the inability to complete the calibration procedure. Task accuracy, completion time and questionnaire responses were complete ($N = 37$). Since an imputation analysis is not recommended with more than 10% missing data (Lodder, 2014), we performed the PCA analysis presented in this article only by considering the 19 participants with a complete dataset.

Time-based data sources were collected using different sampling rates (eye tracking data at 300 Hz and EEG data at 128 Hz). This sampling rate difference should not compromise our statistical outcomes since the measured values, derived from the two data sources, have been aggregated over the whole animation trials as averaged values (e.g., one averaged entropy or alpha power value across all the 16 animated displays for each participant). Other aggregation or interpolation techniques are necessary if, for example, we aimed to measure specific events such as event-related potentials (ERPs). The recorded timestamp of each data source has been used to join and synchronize the different measurements. Both absolute timestamps of eye tracking and EEG data were recorded in milliseconds.

Table 2
Summary of the collected and missing data for the listed data sources, and across expertise groups.

Independent variables	Dependent variables		
Expertise	Response accuracy and time, SSSQ responses and spatial ability scores	EEG data	Eye tracking data
Novices	19	13	12
Experts	18	11	12
All	37	24	24
Missing data	0	13	13

2.7.2. Eye tracking data processing

To capture variability in visual strategies across individuals and expertise levels, we computed first-order (fixation) transition matrix entropies, i.e., transition and stationary entropies (Krejtz et al., 2015). Transition entropy ($\widehat{H_t}$) is computed for individual transition matrices according to Shannon's entropy (Ciuperca & Girardin, 2005; Ekroot & Cover, 1993), which is based on the first-order Markov chain (i.e., the next gaze transition depends only on the present fixation on a certain AOI and not on past eye movement patterns), as shown in equation (2). Our stochastic model of transition matrices relied only on the current participant's visual attention state (and not past states). This is because participant's current state can be viewed as an instance of first-order Hidden Markov Models (HMMs), as suggested by Liechty, Pieters, and Wedel (2003), who used this model to efficiently classify temporal eye movement patterns (i.e., local versus global visual attention states). In our case, however, we were more interested in comparison between transition matrices, as afforded by transition entropy, instead of classification per se.

$$\widehat{H_t} = -\sum_{i\epsilon S}\pi_i \sum_{j\in S} p_{ij} log_2 p_{ij} \tag{2}$$

In this equation, p_{ij} are transition probabilities and π_i stationary probabilities of a participant's AOI switching pattern and by a given sequence $x = (x_0, \ldots, x_n)$. In general, high $\widehat{H_t}$ values indicate that participants frequently switch between the four or eight aircraft depicted on our tested ATC animated displays; thus, their visual search strategies seem to be more exploratory, and their AOI sequence more complex. It can also be indicative of high "expected surprise" (Krejtz et al., 2015) or anxiety (Allsop & Gray, 2014). Conversely, low transition entropy values indicate that participants' eye movements switch less frequently between the depicted moving objects, and are thus indicative of a more focused search behavior, such as highly curious users (Krejtz et al., 2015).

Stationary entropy ($\widehat{H_s}$) is calculated for individual stationary fixation distributions as shown in equation (3):

$$\widehat{H_s} = -\sum_{i\in S}\pi_i log_2 \pi_i \tag{3}$$

Higher $\widehat{H_s}$ values suggest that participants' visual attention is more homogeneously distributed among AOIs, whereas low values indicate that eye fixations are concentrated on a few specific AOIs. Transition matrix entropies were calculated using an R script developed by Krejtz et al. (2015). For further details about how transition and stationary entropy have been computed, see Krejtz et al. (2015).

2.7.3. EEG data processing

EEG raw data were collected using the Emotiv TestBench software (version 2.0, https://www.emotiv.com/, accessed: 21.01.2021). Successively, EEG raw data were processed with EEGLAB (version 14.1.1, https://sccn.ucsd.edu/eeglab/index.php, accessed: 21.01.2021). First, EEG raw data and event information (i.e., type and latency) were imported. Afterwards, continuous EEG data were filtered by applying a short non-linear (IIR) filter (i.e., 0.1–30 Hz), and then EEG artefacts, such as eye blinks or body movements, were removed (i.e., signal segments containing artefacts were first rejected manually and then decomposed by independent component analysis algorithms to eliminate eye blink artefacts). Finally, the filtered data were segmented into epochs, which started 500 ms before and ended 4 s after the onset of each animation. These binned EEG data were then used to calculate average alpha power, FAA, and the engagement index.

Firstly, binned data were used to compute mean alpha (8–12 Hz) band power values for each participant to infer information about participants' cognitive load and determine frontal alpha asymmetry (FAA). Higher frontal alpha asymmetry indicates approach behavior for positive stimuli, but withdrawal behavior from negative stimuli

(Briesemeister et al., 2013). Previous research demonstrates that FAA effects are limited to both frontal electrodes and alpha band (Davidson, Schwartz, Saron, C., Bennett, & Goleman, 1979; Ramsøy, Skov, Christensen, & Stahlhut, 2018; Weinreich, Stephani, & Schubert, 2016). To calculate both alpha power and FAA, we used a custom-made MATLAB graphical user interface called GIVA_EEGtoolbox (available at the following URL: https://gitlab.uzh.ch/giva/geovisense/GIVA_EEGtoolbox, accessed: 21.01.2021). More specifically, alpha band power values were decomposed for each animation segment (i.e., the first 4 s after the start of each animation) using Fourier transform (Tran, Thuraisingham, Wijesuriya, Craig, & Nguyen, 2014) and a 0.5-s Hamming window with overlap of the six prefrontal channels (i.e., F3/F4, F7/F8, FC5/FC6). After baseline correction (i.e., 500 to 0 ms before stimulus onset), we calculated hemispheric asymmetries in the alpha band in the prefrontal brain cortex according to the FAA metric, where FAA values are computed as natural-log transformed difference between the alpha power of right-hemispheric electrodes and their left-hemispheric electrodes (equation (5), Briesemeister et al., 2013):

$$\ln(R) - ln(L)\ ,\ with\ R = \frac{F4 + FC6 + F8}{3}\ and\ L = \frac{F3 + FC5 + F7}{3} \tag{5}$$

In line with Pope et al. (1995), the participants' engagement index was also computed using the GIVA_EEGtoolbox by taking the ratios of beta/(alpha + theta) EEG bandwidths. According to McMahan, Parberry, and Parsons (2015), alpha, beta and theta power ratios of each animation segment (i.e., the first 4 s after the start of each animation) were extracted using a fast Fourier transform (FFT) and a 0.5-s Hamming window with no overlap, averaged for all the 14 EEG channels on the Emotiv headset.

3. Results

First, we fit multilevel models to the data to test our hypotheses about the effects of visual search on task performance. Next, we present outcomes of a Principal Component Analysis (PCA) and related regression analysis to triangulate the variables in this study and highlight the variables that predict task performance (i.e., task accuracy and completion time).

3.1. Descriptive statistics across animations and expertise

Generally, experts completed the task more accurately than novices. However, experts took more time to complete the task overall, especially with continuous animations compared to time-stepped animations, which they were already familiar with (see Fig. 5, see Table 3). Further, novices performed much better with time stepped animations than with continuous animations, and time-stepped animations were completed more accurately overall (see Fig. 5, see Table 4).

3.2. Visual search and task performance: multilevel model results

We utilized a logistic multilevel model appropriate for modeling the effects of binary (0 or 1, incorrect vs correct) outcomes in nested data structures (in this case, trials nested within persons). A logit link function was used to model accuracy outcomes, and a normal link function was used to model the response time outcomes. Prior to fitting the final model, we tested for effects of both types of entropy (transition and stationary), as well as an interaction between animation type and expertise. Since transition entropy did not predict performance, it was left out of the final model. Further, the model failed to converge with an interaction between expertise and animation design, so this was left out of the final model to test our research questions. The final model can be described by Equation (6):

Table 3
Response time across expertise and animations.

	Animation	
Expertise	Time-stepped	Continuous
Experts	M = 49.88 (SD = 8.86)	M = 61.65 (SD = 8.36)
Novices	M = 49.60 (SD = 16.92)	M = 43.95 (SD = 15.57)

Fig. 5. Average percent of response accuracy and time across expertise and animation conditions.

Table 4
Response accuracy across expertise and animations.

	Animation	
Expertise	Time-stepped	Continuous
Experts	M = 86.13 (SD = 15.66)	M = 77.70 (SD = 15.11)
Novices	M = 63.20 (SD = 31.35)	M = 27.78 (SD = 29.55)

$$Y_{ij} = \beta_{00} + \beta_{10}\text{Stationary Entropy}_j + \beta_{01}\text{Expertise}_i + \beta_{02}\ \text{Animation}_i + u_{0j} + r_{ij} \tag{6}$$

where i represents trials and j represents individuals. This model sought to determine if visual search entropy predicts performance, even after controlling for animation design and expertise. We found that higher stationary entropy and expertise result in better accuracy on a given trial (see Table 5, see Fig. 6). While people generally performed better with time-stepped animations, the effect was not statistically significant.

The next model was similar to the first, but with a continuous outcome to model the effects of entropy, animation type, and expertise on response time. In this model, an interaction between animation type and expertise was included. We found that stationary entropy predicts performance even after controlling for animation type and expertise (see Table 6, see Fig. 7). Contrary to our expectations (H1), this suggests that stationary entropy serves as a common visual metric for predicting better task performance, irrespective of training or animation type. Further, there was a significant interaction between expertise and animation type, such that experts took much longer to respond when viewing continuous animations, whereas novices responded in similar times regardless of animation (see Table 6).

3.3. Principal components analysis

To further investigate which variables of this study predict task accuracy and completion time, a Principal Component Analysis (PCA) was conducted. Principal components analysis (PCA) was employed as an exploratory data reduction technique to summarize variables that share common variance into a single component to triangulate the variables in this study. The results of this technique are presented for 19 participants who had a complete dataset. For this reason, these results should be considered as an exploratory cross-validation analysis, and future studies will be required to confirm patterns in our findings.

Prior to conducting the principal components analysis, we utilized a parallel analysis to determine the number of components to extract from our data using the psych package in R (Watkins, 2006). The parallel analysis suggested that one component was present in our data, as the observed eigenvalue on the second component did not exceed the 95th percentile of randomly simulated data. Thus, we extracted one component using predictor variables of interest: spatial abilities, alpha power, frontal alpha asymmetry, engagement index, transition entropy, stationary entropy, and self-reported emotional states (engagement, distress and worry). However, fit was relatively poor (RSMR = 0.20) and several variables did not load highly onto the component.

Table 5
Logistic multilevel model with response accuracy regressed onto stationary entropy, animation type, and expertise.

	Accuracy		
Predictors	Odds Ratios	CI	p
(Intercept)	0.01	0.00–0.76	**0.036**
Stationary entropy	315.78	3.79–26280.34	**0.011**
Expertise	5.93	1.76–19.98	**0.004**
Animation type	0.32	0.10–1.05	0.060
Random Effects			
σ^2	3.29		
$\tau_{00\ \text{subject}}$	1.49		
ICC	0.31		
N_{subject}	23		
Observations	363		
Marginal R^2/Conditional R^2	0.319/0.531		

Therefore, we completed a follow-up principal components analysis using the six variables that loaded highly (greater than or equal to 0.5) with the factor of interest. This resulted in better model fit (*RSMR* = 0.15) and a component comprised of spatial ability, transition entropy, stationary entropy (Fig. 9 shows participants' entropy scores across expertise groups), engagement index and subjective engagement (see Table 7a, Table 7b for full results). Together, these variables explained 60% of variance in the component and were each positively associated with the component. In other words, individuals with higher spatial ability search displays more thoroughly and exhibit higher levels of engagement (measured by EEG and self-reported) compared to individuals with lower spatial ability. This result in part confirms hypothesis H2a - that high gaze entropy coupled with positive emotional states and high spatial skill positively affect task performance.

We then utilized individuals' scores from the principal component analysis to determine if their component score predicted task performance. A linear regression revealed that component scores predicted accuracy ($B = 28.1$, $SE = 4.9$, $p < 0.001$, $R^2 = 0.67$), but not completion time ($B = 3.0$, $SE = 3.8$, $p = 0.43$, $R^2 = 0.04$), see Fig. 8. Using a non-parametric Wilcoxon-Mann-Whitney *t*-test (10,000 sample Monte Carto distribution approximation), we found that experts had higher component scores than novices ($z = 2.12$, $p = 0.03$).

3.4. Effects of engagement on task performance: multilevel model results

Because engagement was important to the component but did not share as much variance with the component, we performed post-hoc analyses to determine if engagement predicted task performance. We ran a similar multilevel model to the model reported in section 3.2. We found that an increase in subjective engagement was associated with a higher likelihood of answering correctly on a given trial, while engagement index (measured by EEG) did not predict meaningful changes in accuracy (see Table 7a,Table7b, see Fig. 10). Further, neither subjective engagement nor engagement index predicted response time (see Table 8).

3.5. Task easiness and enjoyment

Participants' self-assessments concerning task easiness and enjoyment during the experiment revealed that experts judged the task as slightly easier to solve (i.e., 25% easy, 12% neutral, 62% not easy) compared to novices (i.e., 11% easy, 16% neutral, 74% not easy). For self-perceived enjoyment during the experiment, 39% of the experts reported that they enjoyed the experiment (i.e., 39% enjoyed, 28% neutral, 33% did not enjoy), while only 17% of novices enjoyed it (i.e., 17% enjoyed, 28% neutral, 56% did not enjoy).

3.6. Short Stress State questionnaire responses

To gain insight into the viewers' emotional state during the experiment, they reported their subjective judgments about three of the following stress-related emotions collected by the Short Stress State questionnaire (Helton, 2004): their engagement with the task, their distress, and their worry during the experiment (Fig. 11). Although the experts reported greater engagement and less distress and worry than the novices, the levels were not significantly different across expertise groups. However, when we look at their level of engagement across the two animation conditions (Fig. 12), we can see that those experts who solved the task for the condition where the animations familiar (i.e., the time-stepped animations) were significantly more engaged compared to the novices of the same animation condition ($t(16) = -2.12$, $p = 0.05$, $r = 0.47$).

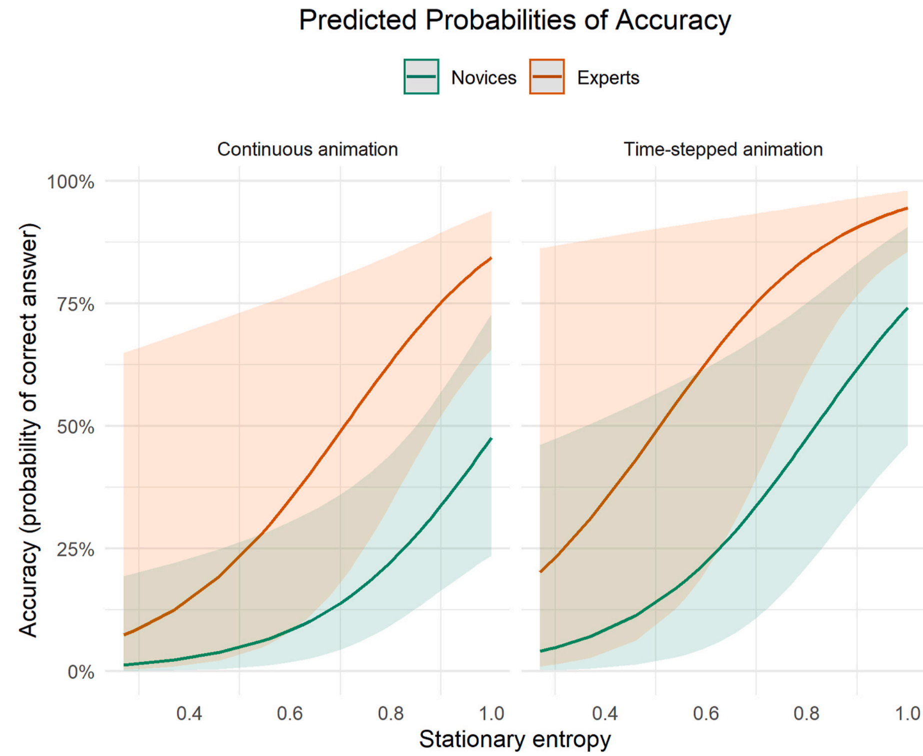

Fig. 6. Predicted response accuracy on a given trial based on animation, expertise, and stationary entropy. Shaded areas correspond to 95% confidence intervals (CIs) of estimated effects.

Table 6
Logistic multilevel model with response time regressed onto stationary entropy, animation type, and expertise.

	Response time		
Predictors	Estimates	CI	p
(Intercept)	18.46	−9.58 – 46.50	0.197
Stationary entropy	38.56	8.47–68.66	**0.012**
Animation type	0.29	−11.07 – 11.66	0.959
Expertise	0.96	−10.56 – 12.48	0.870
Animation*Expertise	32.53	9.74–55.32	**0.005**
Random Effects			
σ^2	509.43		
$\tau_{00subject}$	151.07		
ICC	0.23		
$N_{subject}$	23		
Observations	363		
Marginal R^2/Conditional R^2	0.134/0.332		

4. Discussion

In the current study, we aimed to predict users' performance in an animated ATC detection task using visual search patterns. In addition, we also analyzed the role that viewers' mental and emotional states, and traits play in task performance. To answer our research questions, we utilized a cross-validation approach to analyse participants' visual scanning behavior (with eye tracking), mental engagement (by using electroencephalography; EEG), and self-reported emotional states while scrutinizing various moving aircraft depicted on ATC displays. We measured task performance (response accuracy and completion time) in a real-world ATC visual search task, i.e., detection of task-relevant information in a cluttered animated scene.

4.1. General visual mechanism predicting performance, or constrained by expertise and animation?

Our findings suggest that higher stationary entropy (but not transition entropy) is associated with higher accuracy and shorter completion time, irrespective of training or animation type. Similar to experts, the novices who performed well exhibited higher values of both gaze transition and stationary entropy. As highlighted by Shiferaw et al., 2019, gaze stationary entropy reflects the overall spatial dispersion of gazes on a visual scene, where a decrease of this value may suggest a decline in top-down modulation of the search process. More specifically, a decrease of stationary entropy values, coupled with high transition entropy values, may be associated with top-down interferences that induce the viewers to carefully scrutinize only specific objects that appear on the animated display. In fact, slightly higher transition values of experts, coupled by significant higher stationary entropy, could be an indicator of more homogeneously distributed visual fixations on the visual scene, more gaze transitions or switches between objects (higher exploration modulated by top-down gaze control), and an higher top-down engagement and gaze control compared to novices. Conversely, significant lower stationary entropy, encountered mostly in novices, likely reflects a more intensive attentional focus on specific objects, which in our study are the perceptually salient but task-irrelevant visual elements (i.e., fastest moving aircraft). This is consistent with past studies in aviation (Brams et al., 2018). In contrast, participants showing superior task performance, mostly experts, attended to all the aircraft within the animated scene, including those that were distant from one another, as well as aircraft moving at different speeds. This is in line with previous reports in ATC research (McClung & Kang, 2016; Stein, 1989), which note that ATCo are trained to repeatedly scan the whole display using specific strategies (e.g., by systematically using circular, linear or mixed eye scanning patterns) and at a higher saccadic velocity compared to less trained controllers.

In addition, participants performed the detection task more accurately with the time-stepped animations than with continuous

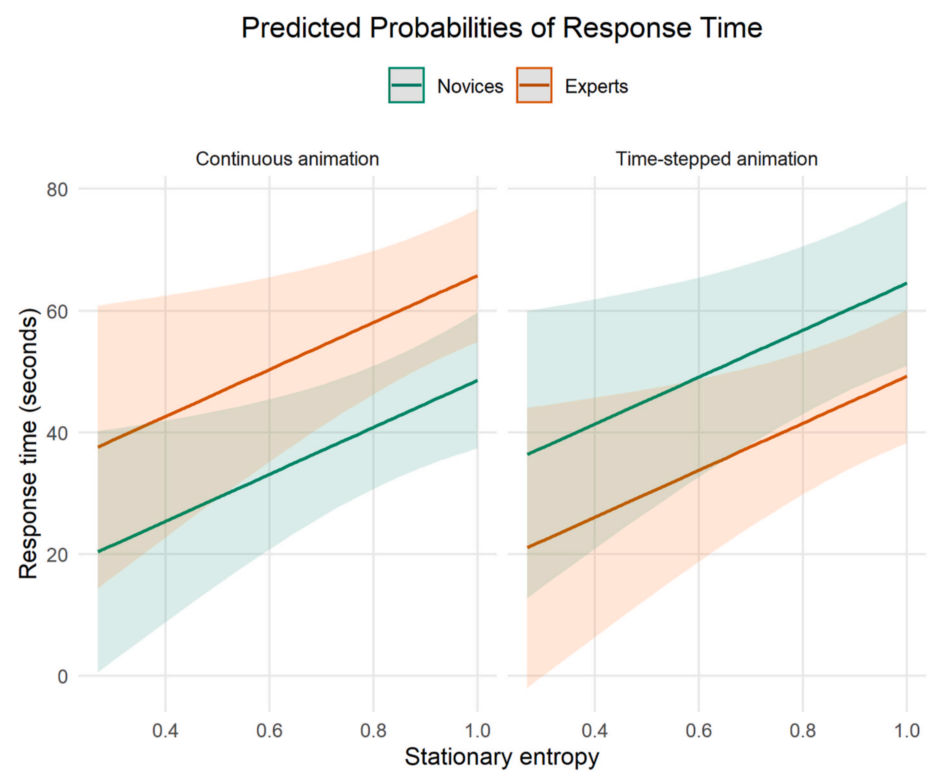

Fig. 7. Predicted response time on a given trial based on animation, expertise, and stationary entropy. Shaded areas correspond to 95% confidence intervals (CIs) of estimated effects.

Table 7a
Results of principal components analysis showing the component consisting of spatial ability, transition entropy, stationary entropy, subjective engagement, and engagement index, including component loadings.

Variable	Loading	Communality
Spatial ability	0.88	0.78
Transition entropy	0.89	0.78
Stationary entropy	0.89	0.80
Engagement index (EEG)	0.53	0.28
Subjective engagement	0.62	0.38

Table 7b
Logistic multilevel model with response accuracy regressed onto engagement index, subjective engagement, animation type, and expertise.

	Accuracy		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	4.49	2.04–9.88	**< 0.001**
Engagement Index	1.01	0.94–1.08	0.827
Subjective Engagement	2.60	0.89–7.57	0.079
Expertise	8.10	2.14–30.67	**0.002**
Animation Design	0.42	0.09–1.86	0.254
Random Effects			
σ^2	3.29		
$\tau_{00\text{subject}}$	1.59		
ICC	0.33		
N_{subject}	22		
Observations	265		
Marginal R^2/Conditional R^2	0.321/0.542		

animations. As partially discussed in Maggi et al. (2016), this is primarily attributable to the performance of novices than those of experts. In fact, experts performed well in both animation conditions, while novices performed much better with time-stepped animations than with continuous animations. The smooth transitions of the frames in the continuous animations have probably accentuated and facilitated the perception of the relative motion between the various visual objects, but has also made more salient those task-irrelevant elements that move at higher speed than the others. This may have misled novices who may have erroneously assigned these aircraft to be the accelerating ones instead of the fastest ones. Contrarily, with continuous animations, those participants who took longer to complete the task likely performed well because they were intensively searching for task-relevant accelerating aircraft. In fact, accelerations are harder to notice and take longer to be detected compared to perceptually salient, but task-irrelevant fast moving aircraft.

4.2. *Which additional individual factors predict performance, and is the contribution of gaze entropy independent from these factors?*

Even though stationary entropy predicted performance independently of spatial ability and engagement (from both neural and subjective judgments), our findings suggest that these individual factors are also related to how users perform visual search. High engaged and spatial ability individuals search a display more completely and fixate for less time on specific moving objects than low engaged and low spatial ability individuals. This is consistent with past findings in aviation (Brams et al., 2018; Diaz-Piedra et al., 2019) as well as other findings suggesting that superior spatial abilities positively influence visuo-spatial task performance (Berney, Bétrancourt, Molinari, & Hoyek, 2015). Across expertise, novices who performed well exhibited an efficient visual scanning strategy (higher levels of transition and stationary entropy), similar to those of experts, as well as high spatial skills. This suggests that higher spatial skills may promote more efficient visual search behaviors in novices, making up for their lack of training with complex animated displays. Contrary to our predictions, cognitive load did not affect task performance.

Contrary to prior work on aviation, which mostly highlighted the negative influence of anxiety or distress on task performance and gaze

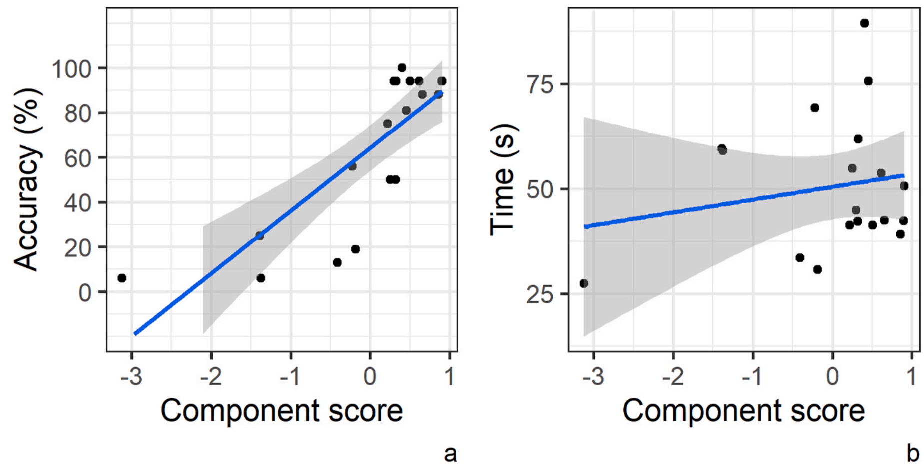

Fig. 8. Response accuracy and time separately regressed onto component score. Component score predicts response accuracy (a, left), but not response time (b, right). Shaded areas correspond to 95% confidence intervals (CIs) of estimated effects.

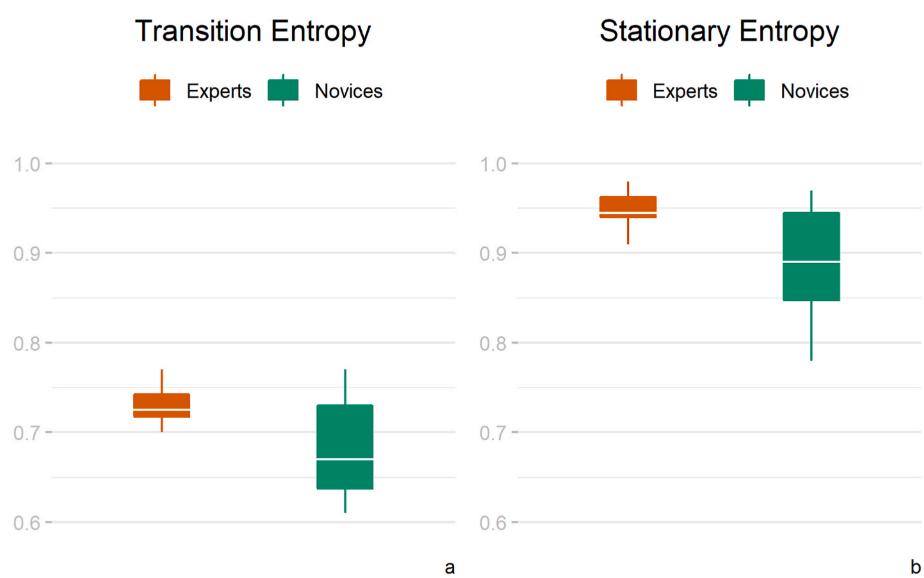

Fig. 9. Participants' transition (a, left) and stationary (b, right) entropy across expertise.

entropy (i.e., increase of anxiety leads to an increase of gaze transitions; Mandrick, Peysakhovich, Rémy, Lepron, & Causse, 2016), we found that especially subjective engagement recorded by self-reported judgments (and not by EEG engagement index) positively affects task accuracy, in both expertise groups and animation types. An increase in the engagement level of the viewer is often linked with an increase in focused attention and alertness (Higgins, 2006; Yerkes & Dodson, 1908). Overall, task performance was then more positively correlated with self-perceived engagement than with distress or worry. However, the pattern differed for experts and novices. Both EEG and participants' subjective judgments indicate that experts were more engaged than novices in solving the required visual search task, but this especially with time-stepped animations. This is not surprising because experts were asked to solve a familiar task relevant to their job with familiar displays; thus, they may have had a higher intrinsic motivation than novices to succeed in the task. Experts' higher mental engagement suggested that they were more focused on the required task than novices. Experts might be closer than novices to an optimal flow state (Csikszentmihalyi, 1990). This result aligns with the inverted-U hypothesis of Yerkes and Dodson (1908) and prior studies in sport research (Khacharem, Zoudji, Kalyuga, & Ripoll, 2013; Saha, Saha, Binti, & Zahir, 2015), which found that people perform best at an optimal level of arousal where their engagement is at the highest level and negative affective states (e.g., distress and worry) are low.

By combining gaze behavior with EEG and self-reports, we thus provide another stepping stone to facilitate the interpretation of behavioral study results, as suggested by Brunyé et al. (2019). Eye tracking data provides useful information for *how* a user interacts with a display: it tells the researcher *where, when,* and for *how long* a user looks at a specific element of the graphic display, but not *why*. Eye tracking studies alone might lead to different observed behavior interpretations, as it is difficult to disentangle cognitive and/or emotional processes when using this data source by itself. Coupled with EEG data and subjective judgment about viewers' emotional states, we are able to see interactions with their cognitive and emotional processes as well as their background: participants' engagement and spatial abilities predict gaze patterns and help to interpret why there are individual task performance differences. Our multivariate analysis served the purpose of simplifying multiple variables into more simple and related components, as well as generate a metric that is useable across a generalizable sample regardless of expertise. This supports the argument that multiple measures should be used when studying complex visuo-spatial decision-making. The measures predict accuracy independently, but also comprise a multifaceted factor that constitutes whether a given individual is

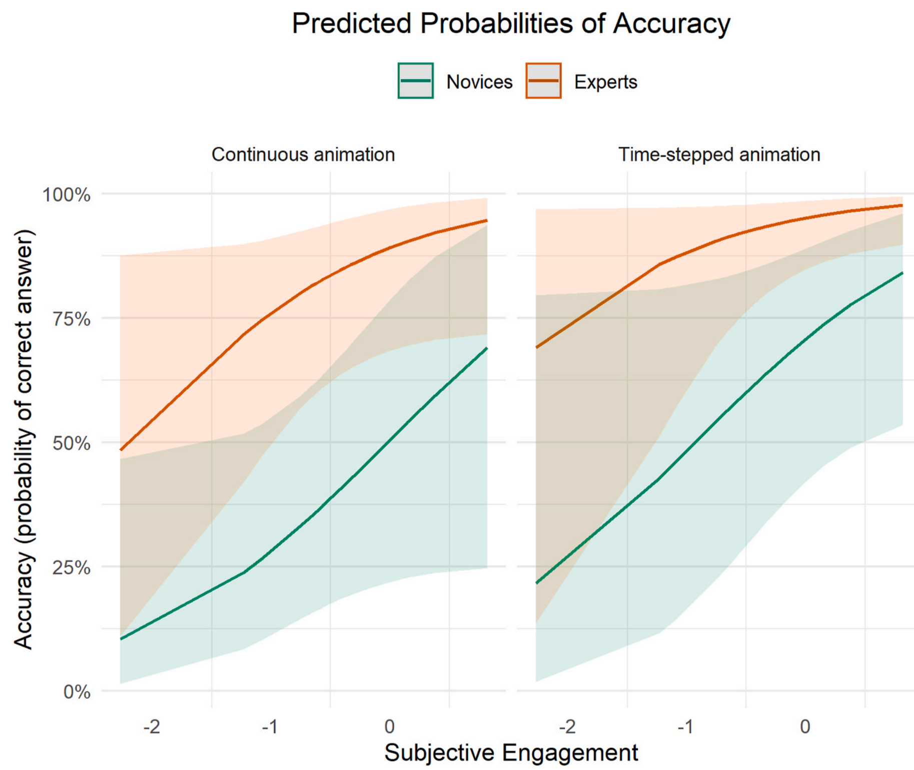

Fig. 10. Predicted response accuracy on a given trial based on animation, expertise, and subjective engagement. Shaded areas correspond to 95% confidence intervals (CIs) of estimated effects.

Table 8
Logistic multilevel model with response time regressed onto engagement index, subjective engagement, animation type, and expertise.

	Response time		
Predictors	Estimates	CI	p
(Intercept)	49.41	43.06–55.75	**< 0.001**
Engagement Index	0.10	−0.32 – 0.52	0.642
Subjective Engagement	−5.95	−15.01 – 3.10	0.198
Expertise	10.22	−1.27 – 21.71	0.081
Animation Design	3.25	−9.66 – 16.16	0.622
Random Effects			
σ^2	540.51		
$\tau_{00subject}$	137.49		
ICC	0.20		
$N_{subject}$	22		
Observations	265		
Marginal R^2/Conditional R^2	0.068/0.257		

successful at completing the task. Our results are consistent with recent work from Padilla, Castro, Quinan, Ruginski, and Creem-Regehr (2019) who demonstrate that traditional measures of accuracy and reaction time should be supplemented by converging measures when assessing the cognitive load of visualization-based decision-making.

5. Future work and conclusions

Ideally, future work will 1) attempt to replicate the presence of the single PCA component we identified, and if found, 2) test whether this component predicts task performance for similar complex visuo-spatial decisions with visualizations, 3) manipulate additional factors increasing cognitive load to more closely simulate real-world decision-making in ATC contexts, either through a concurrent cognitive task, or through introducing other task-relevant information such as the simultaneous processing of auditory and visual stimuli, 4) develop robust conceptual and computational models, to enable effective and efficient coupling of behavioral data streams (e.g., eye tracking with EEG), and 5) further develop general design guidelines for animations to make task-relevant information more salient and animations more engaging. Design guidelines would help enable rapid development of graphic displays that are adapted to specific user characteristics, in particular to both users' cognitive and emotional skills (e.g., prior knowledge)) and to the specific task and usage context (e.g., a stressful situation). For example, to pursue the role of expertise, a potential research objective would be to investigate how to reduce expert response time with design improvements, as experts responded correctly with both animation types, but at the cost of more time. In novices, one could explore those visual variables most likely to draw their attention to the scene elements most relevant to the task, or reduce their focus to those perceptually salient elements that are not relevant to the task.

Our results should be considered a first step to develop a workflow and identify likely areas for future work, given that missing data made our sample relatively small for a PCA and likely underpowered for detecting small effects (Cohen's $f^2 = 0.02$). Although it can be challenging to find task-domain experts for real-world use case experiments, the robust ecological validity of our findings allows guidance of design-based research on critical decision-making displays. Furthermore, the assessment of animation designs might be influenced by other user characteristics. Beyond group differences in expertise and individual differences in cognitive and emotional abilities, participants' age might affect task performance (Guest, Howard, Brown, & Gleeson, 2015). Given that some cognitive abilities are known to decline after 20–30 years of age (Driscoll, Hamilton, Yeo, Brooks, & Sutherland, 2005; Salthouse, 2009), participants' age differences should be further investigated in future studies on visuo-spatial information displays. In addition, it would be helpful to know how our findings might be extended to different usage contexts, such as in private and public transportation systems, sport analytics, movement ecology, human health, and pedestrian navigation (Dodge et al., 2016), where dynamic visualizations are used for data analysis and decision making. This would allow applied

Fig. 11. SSSQ's self-perceived engagement, distress and worry across expertise for all participants (N = 37).

Fig. 12. SSSQ's subjective engagement across expertise and animation design for all participants (N = 37).

researchers to develop robust design guidelines for animations to facilitate decision making with a broad range of dynamic displays and users. Our study serves as a first step towards that goal by identifying a common visual mechanism - stationary gaze entropy - and other individual differences factors - spatial abilities and engagement - that directly contribute to performance in object detection with dynamic displays irrespective of animation type and training.

Ethics approval and consent to participate

This study was run in accordance with the recommendations of the Temple University's Human Research Protection Program (HRPP). The ethics procedure was approved by Temple University's Office for Human Subjects Protections, Institutional Review Board (IRB) prior to data collection.

Consent for publication

All participants, both ATC experts and ATC novices, gave written informed consent to participate in this study.

Availability of data and materials

The statistical data, R-code and animations (inclusive Processing codes) that support the findings of this study are available in OSF at https://osf.io/d2jpg/. If you have any questions related to the data, please contact the corresponding author, Sara Lanini-Maggi at sara.maggi@geo.uzh.ch. The MATLAB code of our EEG tool is available at the following URL: https://gitlab.uzh.ch/giva/geovisense/GIVA_EEGtoolbox (accessed: 21.01.2021).

Funding

This work was supported by the European Research Council (ERC) Advanced Grant GeoViSense [#740426] and the Digital Society Initiative of the University of Zurich.

Authors' contributions

S.L.-M. took the lead in writing this manuscript, as well as conceived, designed and performed the data analyses, together with I.T.R., and experiment, and collected the data, presented in this manuscript. I.T.R. not only performed part of the statistical analyses, but also contributed substantially in writing the manuscript. T.E.S. contributed substantially in the conception and design of the analyses and in providing critical feedback to the manuscript and statistical data analyses. C.H. helped S. L.-M. in designing, planning and carrying out the experiment with air traffic controllers at ENAC, Toulouse (FR). C.H. also provided critical feedback to the manuscript, especially in those parts related to real-world air traffic control tasks and scenarios. A.T.D. contributed in developing and making available to S.L.-M. the used tools to analyse the collected eye tracking data, as well as in providing critical feedback to the manuscript, especially in those parts related to eye movement analyses and results. B.B.B. making available to S.L.-M. the used tools to analyse the collected EEG data, as well as in providing critical feedback to the manuscript, especially in those parts related to EEG data analyses and results. J.L. in developing and making available to S.L.-M. the used tools to analyse the collected EEG data, as well as in providing critical feedback to the manuscript, especially in those parts related to EEG data analyses and results. S.I.F. is the project leader of the GeoViSense project and, together with S.L.-M., conceived and gave substantial support in designing and supervising the presented user study, and manuscript.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgements

We would like to express our gratitude to the air traffic controllers and students who participated in our user study. We are also grateful for Dr. Jean-Paul Imbert's and Dr. Maxime Cordeil's invaluable support during the design phase of the experiment and for the programming of the animated displays. We are also particularly grateful to Dr. Tyler Thrash and Bingjie Cheng for their critical feedback on earlier drafts of this manuscript. Thanks to Dr. Jennifer Bartmess for proofreading this manuscript.

References

Allsop, J., & Gray, R. (2014). Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *Journal of Applied Research in Memory and Cognition, 3*(2), 63–71. https://doi.org/10.1016/j.jarmac.2014.04.010.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., et al. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation Space & Environmental Medicine, 78*(5 Suppl). B231—44. Retrieved from http://europepmc.org/abstract/MED/17547324.

Berney, S., Bétrancourt, M. B., Molinari, G., & Hoyek, N. (2015). How spatial abilities and dynamic visualizations interplay when learning functional anatomy with 3D anatomical models. *Anatomical Sciences Education*. https://doi.org/10.1002/ase.1524.

Brams, S., Hooge, I. T. C., Ziv, G., Dauwe, S., Evens, K., De Wolf, T., et al. (2018). Does effective gaze behavior lead to enhanced performance in a complex error-detection cockpit task? *PloS One, 13*(11), Article e0207439. https://doi.org/10.1371/journal.pone.0207439.

Briesemeister, B. B., Tamm, S., Heine, A., & Jacobs, A. M. (2013). Approach the good, withdraw from the bad — a review on frontal alpha asymmetry measures in applied psychological research. *Psychology, 4*(3), 261–267. https://doi.org/10.4236/psych.2013.43A039.

Brunyé, T. T., Drew, T., Weaver, D. L., & Elmore, J. G. (2019, December 1). A review of eye tracking for understanding and improving diagnostic interpretation. *Cognitive Research: Principles and Implications. Springer*. https://doi.org/10.1186/s41235-019-0159-2.

Bryman, A. (1984). The debate about quantitative and qualitative research: A question of method or epistemology? *British Journal of Sociology, 35*(1). https://doi.org/10.2307/590553, 75.

Bunch, R. L., & Lloyd, R. E. (2006). The cognitive load of geographic information. *The Professional Geographer, 58*(2), 209–220. https://doi.org/10.1111/j.1467-9272.2006.00527.x.

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences, 9*(7), 349–354. https://doi.org/10.1016/j.tics.2005.05.009.

Chevalier, F., Riche, N. H., Plaisant, C., Chalbi, A., Hurter, C., & An, C. H. (2016). Animations 25 years later: New roles and opportunities. In *In AVI 16, international working conference on advanced visual interfaces, ACM, jun 2016, bari, Italy*. https://doi.org/10.1145/2909132.2909255ï, 280.

Ciuperca, G., & Girardin, V. (2005). On the estimation of the entropy rate of finite Markov chains. In *Conference "international symposium on applied stochastic models and data analysis" (ASMDA 2005), may, 17, 18, 19, 20, 2005. Brest (France)*.

Coan, J. A., & Allen, J. J. (2004). Frontal EEG asymmetry as a moderator and mediator of emotion. *Biological Psychology, 67*(1–2), 7–50. https://doi.org/10.1016/j.biopsycho.2004.03.002.

Cordeil, M., Dwyer, T., & Hurter, C. (2016). Immersive solutions for future air traffic control and management. In *Proceedings of the 2016 ACM companion on interactive surfaces and spaces - ISS companion '16* (pp. 25–31). New York, New York, USA: ACM Press. https://doi.org/10.1145/3009939.3009944.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.

Dalrymple, K. A., Manner, M. D., Harmelink, K. A., Teska, E. P., & Elison, J. T. (2018). An examination of recording accuracy and precision from eye tracking data from toddlerhood to adulthood. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.00803.

Davidson, R. J., Schwartz, G. E., Saron, C., Bennett, J., & Goleman, D. J. (1979). Frontal versus parietal EEG asymmetry during positive and negative affect. *Psychophysiology, 16*, 202–203.

Derryberry, D., & Tucker, D. M. (1994). Motivating the focus of attention. In *The heart's eye* (pp. 167–196). https://doi.org/10.1016/B978-0-12-410560-7.50014-4. Elsevier.

Diaz-Piedra, C., Rieiro, H., Cherino, A., Fuentes, L. J., Catena, A., Stasi, Di, et al. (2019). The effects of flight complexity on gaze entropy: An experimental study with fighter pilots. *Applied Ergonomics, 77*, 92–99. https://doi.org/10.1016/j.apergo.2019.01.012.

DiBiase, D., MacEachren, A. M., Krygier, J. B., & Reeves, C. (1992). Animation and the role of map design in scientific visualization. *Cartography and Geographic Information Systems, 19*(4), 201–214. https://doi.org/10.1559/152304092783721295.

Dodge, S., Weibel, R., Ahearn, S. C., Buchin, M., & Miller, J. A. (2016). Analysis of movement data. *International Journal of Geographical Information Science, 30*(5), 825–834. https://doi.org/10.1080/13658816.2015.1132424.

Drew, T., Le-Hoa Vo, M., Olwal, A., Jacobson, F., Seltzer, S. E., & Wolfe, J. M. (2013). Scanners and drillers: Characterizing expert visual search through volumetric images. *Journal of Vision, 13*(10). https://doi.org/10.1167/13.10.3.

Driscoll, I., Hamilton, D. A., Yeo, R. A., Brooks, W. M., & Sutherland, R. J. (2005). Virtual navigation in humans: The impact of age, sex, and hormones on place learning. *Hormones and Behavior, 47*(3), 326–335. https://doi.org/10.1016/j.yhbeh.2004.11.013.

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers, 34*(4), 455–470. https://doi.org/10.3758/BF03195475.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66*(3), 183–201. https://doi.org/10.1037/h0047707.

Ekroot, L., & Cover, T. M. (1993). The entropy of markov trajectories. *IEEE Transactions on Information Theory, 39*(4), 1418–1421. https://doi.org/10.1109/18.243461.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). Manual for the kit of factor-referenced cognitive tests. Princeton N.J. ;: Education Testing Service. Retrieved from https://www.ets.org/Media/Research/pdf/Manual_for_Kit_of_Factor-Referenced_Cognitive_Tests.pdf.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(1), 32–64. https://doi.org/10.1518/001872095779049543.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007, May). Anxiety and cognitive performance: Attentional control theory. *Emotion*. https://doi.org/10.1037/1528-3542.7.2.336.

Fabrikant, S. I., Rebich, S., Montello, D. R., Andrienko, G., & Andrienko, N. (2008). A visual analytics approach to evaluate inference affordance from animated map displays. In *In proceeding of fifth international conference on geographic information science: Pre-conference workshop on geospatial visual analytics (pp. 1–7). Park city, Utah*. https://doi.org/10.5167/uzh-10198.

Fish, C. S. (2015). Cartographic challenges in animated mapping. In *In ICA workshop on envisioning the future of cartographic research at the international cartographic conference. Curitiba, Brazil*. Retrieved from https://cogvis.icaci.org/pdf/icc2015/Fish_ICA_Curitiba.pdf.

Fox, J., Merwin, D., Marsh, R., McConkie, G., & Kramer, A. (1996). Information extraction during instrument flight: An evaluation of the validity of the eye-mind hypothesis. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting, 40*(2), 77–81. https://doi.org/10.1177/154193129604000215.

Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013, March). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2013.01.010.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics, 24* (6), 631–645. https://doi.org/10.1016/S0169-8141(98)00068-7.
Guest, D., Howard, C. J., Brown, L. A., & Gleeson, H. (2015). Aging and the rate of visual information processing. *Journal of Vision, 15(14), 10*. https://doi.org/10.1167/15.14.10.
Harmon-Jones, E., Gable, P. A., & Price, T. F. (2012). The influence of affective states varying in motivational intensity on cognitive scope. *Frontiers in Integrative Neuroscience, 6*(73). https://doi.org/10.3389/fnint.2012.00073.
Heger, D., Putze, F., Schultz, T., & n R. (2010). Online workload recognition from EEG data during cognitive tests and human-computer interaction. In S. T. Dillman, J. Beyerer, & U. D. Hanebeck (Eds.), *KI 2010: Advances in artificial intelligence (KI 2010: A* (pp. 410–417). Berlin, Heidelberg: Springer. Retrieved from https://www.csl.uni-bremen.de/cms/images/documents/publications/HegerPutzeSchultz_KI10.pdf.
Helton, W. S. (2004). Validation of a short stress state questionnaire. In *, Vol. 48. Proceedings of the human factors and ergonomics society annual meeting* (pp. 1238–1242). Los Angeles, CA: SAGE PublicationsSage CA. https://doi.org/10.1177/154193120404801107.
Higgins, E. T. (2006). Value from hedonic experience and engagement. *Psychological Review, 113*(2), 439–460. https://doi.org/10.1037/0033-295X.113.3.439.
Holmqvist, K. (2011). Eye tracking: A comprehensive guide to methods and measures. Oxford university press. Retrieved from https://global.oup.com/academic/product/eye-tracking-9780199697083?cc=ch&lang=en&.
Hurter, C., & Conversy, S. (2008). Towards characterizing visualizations. In *Interactive systems. Design, specification, and verification* (pp. 287–293). https://doi.org/10.1007/978-3-540-70569-7_26. Berlin, Heidelberg: Springer Berlin Heidelberg.
Hurter, C., Conversy, S., & Vinot, J.-L. (2009). Temporal data visualizations for air traffic controllers (ATC). *Chips*, 2009. Retrieved from https://hal.archives-ouvertes.fr/hal-00879033.
Hurter, C., Lesbordes, R., Letondal, C., Vinot, J.-L., & Conversy, S. (2012). Strip-TIC: Exploring augmented paper strips for air traffic controllers. In *In AVI 2012, international working conference on advanced visual interfaces, may 2012, capri island, Italy* (pp. 225–232). https://doi.org/10.1145/2254556.2254598ï.
Hyrskykari, A., Ovaska, S., Majaranta, P., Räihä, K.-J., & Lehtinen, M. (2008). Gaze path stimulation in retrospective think-aloud. *Journal of Eye Movement Research, 2*(4), 1–18. https://doi.org/10.16910/jemr.2.4.5.
Jarodzka, H., & Gerjets, P. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction, 20*(2), 146–154. https://doi.org/10.1016/J.LEARNINSTRUC.2009.02.019.
Kamzanova, A. T., Kustubayeva, A. M., & Matthews, G. (2014). Use of EEG workload indices for diagnostic monitoring of vigilance decrement. *Human Factors, 56*(6), 1136–1149. https://doi.org/10.1177/0018720814526617.
Khacharem, A., Zoudji, B., Kalyuga, S., & Ripoll, H. (2013). The expertise reversal effect for sequential presentation in dynamic soccer visualizations. *Journal of Sport & Exercise Psychology, 35*(3), 260–269. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23798589.
Klein, T., van der Zwan, M., & Telea, A. (2014). Dynamic multiscale visualization of flight data. In *International conference on computer vision theory and applications (VISAPP)* (pp. 104–114).
de Koning, B. B., Jarodzka, H. (2017). Attention guidance strategies for supporting learning from dynamic visualizations. In Learning from Dynamic Visualization (pp. 255–278. *Cham: Springer International Publishing.* https://doi.org/10.1007/978-3-319-56204-9_11.
Kothe, C. A., & Makeig, S. (2011). Estimation of task workload from EEG data: New and current tools and perspectives. In *2011 annual international conference of the IEEE engineering in medicine and biology society* (pp. 6547–6551). https://doi.org/10.1109/IEMBS.2011.6091615. IEEE.
Krejtz, K., Duchowski, A. T., & Krejtz, I. (2014). Entropy-based statistical analysis of eye movement transitions. ETRA. Retrieved from http://en.
Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., González Perilli, F., Pires, A., et al. (2015). Gaze transition entropy. *ACM Transactions on Applied Perception, 13*(1), 1–20. https://doi.org/10.1145/2834121.
Kriglstein, S., Pohl, M., & Stachl, C. (2012). Animation for time-oriented data: An overview of empirical research. In *2012 16th international conference on information visualisation* (pp. 30–35). https://doi.org/10.1109/IV.2012.16. IEEE.
Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. *International Journal of Human-Computer Studies, 65*(11), 911–930. https://doi.org/10.1016/j.ijhcs.2007.06.005.
Letondal, C., Hurter, C., Lesbordes, R., Vinot, J.-L., & Conversy, S. (2013). Flights in my hands. In *Proceedings of the SIGCHI conference on human factors in computing systems - CHI '13* (p. 2175). New York, New York, USA: ACM Press. https://doi.org/10.1145/2470654.2481300.
Liechty, J., Pieters, R., & Wedel, M. (2003). Global and local covert visual attention: Evidence from a Bayesian hidden Markov model. *Psychometrika, 68*(4), 519–541. https://doi.org/10.1007/BF02295608.
Lodder, P., & G. J. M.. (2014). To impute or not impute: That's the question. In H. J. Adèr (Ed.), *Advising on research methods. Johannes van Kessel Publishing.* Retrieved from https://www.paultwin.com/wp-content/uploads/Lodder_1140873_Paper_Imputation.pdf.
Lowe, R., & Schnotz, W. (2008). *Learning with animation: Research implications for design.* Cambridge University Press.
MacKay, W. E. (1999). Is paper safer? The role of paper flight strips in air traffic control. *ACM Transactions on Computer-Human Interaction, 6*(4), 311–340. https://doi.org/10.1145/331490.331491.
Mackay, W. E., Fayard, A.-L., Frobert, L., & Medini, L. (1998). Reinventing the familiar: Exploring an augmented reality design space for air traffic control. In *In proceedings of the 1998 conference on human factors in computing systems, CHI - los angeles, CA, USA* (pp. 558–565). ACM. Retrieved from https://nyuscholars.nyu.edu/en/publications/reinventing-the-familiar-exploring-an-augmented-reality-design-sp.
Maggi, S., Fabrikant, S. I., Imbert, J., & Hurter, C. (2016). How do display design and user characteristics matter in animations? An empirical study with air traffic control displays. *Cartographica, 51*(1), 25–37. https://doi.org/10.3138/cart.51.1.3176.
Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., & Causse, M. (2016). Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biological Psychology, 121*, 62–73. https://doi.org/10.1016/j.biopsycho.2016.10.002.
Mayer, R. E. (2012). Cognitive theory of multimedia learning. In *The cambridge handbook of multimedia learning* (pp. 31–48). Cambridge University Press. https://doi.org/10.1017/cbo9780511816819.004.
McClung, S. N., & Kang, Z. (2016). Characterization of visual scanning patterns in air traffic control. *Computational Intelligence and Neuroscience, 2016*, 1–17. https://doi.org/10.1155/2016/8343842.
van de Merwe, K., van Dijk, H., & Zon, R. (2012). Eye movements as an indicator of situation awareness in a flight simulator experiment. *The International Journal of Aviation Psychology, 22*(1), 78–95. https://doi.org/10.1080/10508414.2012.635129.
McMahan, T., Parberry, I., & Parsons, T. (2015). Evaluating player task engagement and arousal using electroencephalography. *Procedia Manuf., 3*, 2303–2310. https://doi.org/10.1016/j.promfg.2015.07.376.
Padilla, L. M. K., Castro, S. C., Quinan, P. S., Ruginski, I. T., & Creem-Regehr, S. H. (2019). Toward objective evaluation of working memory in visualizations: A casestudy using pupillometry and a dual-task paradigm. *IEEE Transactions on Visualization and Computer Graphics*. https://doi.org/10.1109/tvcg.2019.2934286, 1–1.
Pfeiffer, L., Valtin, G., Müller, N. H., & Rosenthal, P. (2016). The mental organization of air traffic and its implications to an emotion sensitive assistance system. *International Journal on Advances in Life Sciences, 8*(1&2).
Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology, 40*(1–2), 187–195. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7647180.
Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*(1), 3–25. https://doi.org/10.1080/00335558008248231.
Ramsøy, T. Z., Skov, M., Christensen, M. K., & Stahlhut, C. (2018). Frontal brain asymmetry and willingness to pay. *Frontiers in Neuroscience, 12*. https://doi.org/10.3389/fnins.2018.00138. MAR.
Robertson, G., Fernandez, R., Fisher, D., Lee, B., & Stasko, J. (2008). Effectiveness of animation in trend visualization. *IEEE Transactions on Visualization and Computer Graphics, 14*, 1325–1332. https://doi.org/10.1109/TVCG.2008.125.
Roth, R. E. (2013). Interactive maps: What we know and what we need to know. *Journal of Spatial Information Science*. https://doi.org/10.5311/JOSIS.2013.6.105.
Saha, S., Saha, S., Binti, E., & Zahir, M. (2015). Psychobiological indices of emotionality as predictor of reaction ability in high sport performance. *International Medical Journal, 22*.
Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging, 30*(4), 507–514. https://doi.org/10.1016/j.neurobiolaging.2008.09.023.
Schinazi, V. R., & Thrash, T. (2018). Cognitive neuroscience of spatial and geographic thinking. In D. R. Montello (Ed.), *Handbook of behavioral and cognitive Geography* (pp. 154–174). Edward Elgar Publishing. https://doi.org/10.4337/9781784717544.00016.
Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 379–423.
Shiferaw, B., Downey, L., & Crewther, D. (2019, January 1). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience and Biobehavioral Reviews. Elsevier Ltd.* https://doi.org/10.1016/j.neubiorev.2018.12.007.
Shipley, T. F., Fabrikant, S. I., & Lautenschütz, A.-K. (2013). Creating perceptually salient animated displays of spatiotemporal coordination in events. In M. Raubal, D. M. Mark, & A. U. Frank (Eds.), *Cognitive and linguistic aspects of geographic space. New perspectives on geographic information research* (pp. 259–270). Heidelberg: Springer. https://doi.org/10.1007/978-3-642-34359-9_14.
Stein, E. S. (1989). Air traffic controller scanning and eye movements in search of information - a literature review. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a206709.pdf.
Tran, Y., Thuraisingham, R., Wijesuriya, N., Craig, A., & Nguyen, H. (2014). Using S-transform in EEG analysis for measuring an alert versus mental fatigue state. In *2014 36th annual international conference of the IEEE engineering in medicine and biology society, EMBC 2014* (pp. 5880–5883). https://doi.org/10.1109/EMBC.2014.6944966. Institute of Electrical and Electronics Engineers Inc.
Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57*, 247–262. https://doi.org/10.1006/ijhc.1017.
Wass, S. V., Forssman, L., & Leppänen, J. (2014). Robustness and precision: How data quality may influence key dependent variables in infant eye-tracker analyses. *Infancy, 19*(5), 427–460. https://doi.org/10.1111/infa.12055.
Watkins, M. W. (2006). Determining parallel analysis criteria. *Journal of Modern Applied Statistical Methods, 5*(2), 344–346. https://doi.org/10.22237/jmasm/1162354020.
Weinreich, A., Stephani, T., & Schubert, T. (2016). Emotion effects within frontal alpha oscillation in a picture oddball paradigm. *International Journal of Psychophysiology, 110*, 200–206. https://doi.org/10.1016/j.ijpsycho.2016.07.517.
Westerman, S. J., Sutherland, E. J., Robinson, L., Powell, H., & Tuck, G. (2007). A multi-method approach to the assessment of web page designs. In *Affective computing and*

intelligent interaction (pp. 302–313). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74889-2_27.

Wickens, C., & Tsang, P. S. (2015). Workload. *APA Handbook of Human Systems Integration.*, 277–292. https://doi.org/10.1037/14528-018.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-017-0058.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*(5), 459–482. https://doi.org/10.1002/cne.920180503.